Should a 'Natural Monopolist' Be Subject to Competition?  
With Special Reference to Cellular Mobile Telephone Services in Australia

Yew-Kwang Ng*  
Department of Economics  
Monash University

Abstract

The argument against competition based on natural monopoly is questioned. A true natural monopolist needs no protection. Rapid product innovation and technical advance render the telecommunications industry not a natural monopolist.

Telephones may involve two sources of external benefits: benefits to callers and benefits to receivers. Neither source justifies treating country areas more favourably since the access externality also applies, probably with larger magnitudes, to metropolitan areas.

The Cellular Mobile Telephone Service (CMTS) involves the use of parts of the radio frequency spectrum. The spectrum should be treated as a scarce resource. Competitive bidding is probably the most efficient method of allocation if the whole spectrum is to be reallocated. However, with only a small band allocated to CMTS, auctioning need not be the most efficient outcome though it is better than maintaining its monopolised usage.

Even if competition results in a higher cost initially, the gain in dynamic efficiency can offset this initial loss. A minimum estimate of the benefits to the public of introducing competition in CMTS gives a present value of $2.4 billion.

1. Introduction

Consistent with the global movements towards liberalisation and privatisation, there have been significant changes towards deregulation in Australia in recent years. AUSTEL (Australian Telecommunications Authority) was established to ensure, among others, that competitive activities are carried out on a 'level playing field'. In August 1989, Mrs Ros Kelly, Minister for Telecommunications and Aviation Support, issued ministerial guidelines asking AUSTEL to review whether an additional operator in cellular mobile telephone services (CMTS) should be licensed to end Telecom's monopoly.

While this article is written with special reference to CMTS, most of the analysis is relevant to the more general question of whether a natural monopolist should be exempt from competition. In fact, our conclusion applies, a fortiori, to non-natural monopolists.

A natural monopolist is one where a single producer can supply the whole market demand at least cost due to scale economies. In the absence of scale economies, virtually everyone is in favour of competition. For a natural monopoly, opinions are less uniform. Section 2 of this article argues in favour of permitting potential competition and Sections 3, 4 and 5 examine the specific case of Australian telecommunications with special reference to CMTS.

2. Scale Economies and Contestability

2.1 Scale Economies and Natural Monopolies

It is not denied that, in the presence of significant scale economies, an industry may best have only one producer. Rather, our argument in favour of liberalisation runs as follows. First, if a monopoly is really natural due to significant scale economies, it does not need protection or regulation prohibiting competition for it to remain the sole supplier in the market. Secondly, the existence of significant scale economies in the long run may be questioned. Even if they exist, their significance is likely to be overwhelmed by the long-run benefits of competition in promoting quality improvements, product innovation and technical advance.

For simplicity, consider first a static case of a single and homogeneous product. As illustrated in Figure 1, the average (that is, per unit) cost curve AC is downward sloping, reflecting scale economies (or economies of outlay). If the existing monopolist prices its output at \( p^1 \) as determined by the intersection of the AC curve with the demand curve D, it will be able to:

(i) satisfy demand at the price charged;

(ii) make no loss; and

![Figure 1](image-url)
Should a 'Natural Monopolist' Be Subject to Competition?

(iii) prevent any profitable entry of a firm with no lower cost of production than itself.

This is the case of a sustainable monopoly (first defined by Baumol, Bailey and Willig 1977, and Panzar and Willig 1977). No protection or regulation is needed for the monopolist to remain a single supplier.

It is true that there are cases where:

(a) a single producer can supply the market demand at a lower total cost than two or more producers;

(b) a monopolist may not be sustainable in the sense of meeting conditions (i) to (iii) above.

This is illustrated in Figure 2 (see, for example, Sharkey 1982, p. 89). To ensure that entry is unattractive, the monopolist must set a price no higher than $p_2$. But to cover costs as well as satisfying demand, it must set a price no lower than $p_3$. Since $p_3 > p_2$, this is impossible.

Note that the above illustration on an unsustainable monopolist relies on the fact that the AC curve is upward sloping before its intersection with the demand curve. This means that, at least at the margin, there are no economies of scale but rather diseconomies of scale. It may still be true that, to produce the output $Q_3$, one firm is less costly than two firms. However, given a substantial upward-sloping section in the AC curve, the cost saving of just one firm is unlikely to be big, if any. In the case of Figure 2, it is twice the shaded area. In the long run, this small saving is likely to be overwhelmed by the dynamic benefits of competition.

It might be thought that the cost saving could be quite big if the AC curve was the $AC'$ curve. However, such an average cost curve involves a negative marginal cost at some range and does not apply in practice.

In fact, a natural monopolist is likely to be able to do considerably better than illustrated in Figure 1. This is so since the entrant cannot assume that it can replace the existing monopolist as the single producer. (In technical terms, the preceding argument depends on the questionable Bertrand-Nash assumption that the incumbent will keep its price unchanged so that the entrant can capture the whole market by a marginally lower price. See Brock 1983.) Rather, should it enter, it will most probably become a duopolist (before considering any further entry). A most favourable assumption for the entrant is that it will share the market 50–50 with the first producer.

With the simplifying assumption of a homogeneous product, we have to assume that the two firms would have to adopt the same price since otherwise the higher priced one will lose all the market. Each firm expects to face the demand curve $D/2$; that is, half of the total market demand. If the AC curve lies entirely above $D/2$, the existing monopolist may be able to adopt any price to maximise profit without inviting entry.

If the half-market demand curve $D/2$ lies above some section of the common AC curve, the monopolist may be constrained by the threat of entry to price no higher than $p_4$. If this is the case, it is making a handsome profit even at a price slightly lower than $p_4$.

It may be thought that, expecting to operate on the demand curve $D/2$, the prospective entrant may enter even if the existing monopolist prices as low as $p_5$, since its entrance may force the first firm to increase its price to a level above $p_4$. This is extremely unlikely though it is a game-strategic situation without a definite solution. Provided that the incumbent sticks to the price of $p_5$, the entrant faces big losses on entry. Given the real-world superiority of the incumbent in capital investment committed and clientele, it is almost certain that the entrant will have a hopeless battle. It is likely that, for a situation as illustrated in Figure 3, the existing monopolist may be able to charge any price between $p_5$ and $p_4$ without inviting entry provided that it stands ready to lower the price to $p_4$ upon entry.
Should a 'Natural Monopolist' Be Subject to Competition?

Given the advantages enjoyed by the existing producer, the assumption that the entrant can enjoy half of the market upon entry must be relaxed. The entrant has to build up from scratch. In the case of the British duopoly in telecommunications, the second firm (Mercury) is still only a few per cent the size of the first firm (BT). If competition is inefficient due to scale economies, it is unlikely that the existing natural monopolist needs any protection or regulation to maintain the monopoly. Rather, the reverse situation where competition is desirable but prevented from taking place is more likely to apply. These conclusions are consistent with those of Hartley and Trengove (1987, pp. 42, 51–2).

2.2 The Efficiency of Contestable Markets

Allowing a second producer to challenge the monopoly makes the market more contestable. Due to the heavy sunk costs element in the telecommunications industry, it is by no means a perfectly contestable market. Nevertheless, just as perfect competition provides insights into the efficiency of (not necessarily perfect) competition, the theory of perfect contestability also throws light on the relevant issues.

A perfectly contestable market is one in which entry is free and exit is costless. (See Baumol 1982, p. 3.) Freedom of entry not only means the absence of any restriction on entry but also the absence of any cost and demand disadvantages. The requirement of costless exit rules out any sunk costs. Note that this differs from fixed costs. Fixed costs need not be sunk even in the short run. For example, suppose a firm needs a factory of a given size before the first unit of output can be produced; these costs are fixed. If the firm bought the factory for $10 million but could only resell it for $7 million, then $3 million were sunk. But if it could resell it for $10 million, then no costs were sunk: see Baumol (1982, p. 4).

The major efficiency implications of a perfectly contestable market are summarised below.

(i) There must be an absence of any inefficiency in production in industry equilibrium.

(ii) Zero profits must characterise any equilibrium.

(iii) No product can be sold at a price less than its marginal cost; with two or more sellers, their prices (same across different sellers) must equal marginal costs.

(iv) Two firms may be enough to guarantee optimality.

The theory suggests that what is important for efficiency may not be the number of firms actually in production but rather the freedom of entry and exit. The relaxation of the unrealistic assumption of perfect contestability makes the existing producer able to enjoy supernormal profits and prevent entry.

While the theory of contestability as developed by Baumol, Panzar, Willig and others is very much supportive of competition, there is one specific argument by Baumol that is contrary to the general tendency. This is the argument that 'where there are economies of scale in the production of durable capital, ... the least costly producer is in the long run vulnerable to entry or replacement by rivals whose appearance is inefficient because it wastes valuable social resources' (Baumol 1982, p. 3). However, this conclusion ignores the required revaluation of capital and assumes significant sunk costs, contrary to the requirement of contestability.

2.3 Dynamic Efficiency

The preceding argument assumes a homogeneous product. In practice, a new producer probably offers a differentiated product providing more scope for consumers' choice. Thus, even if market supply does not increase, the benefits enjoyed by consumers are increased. Unless aggregate costs increase by more than the increase in aggregate benefits, the change is desirable.

Aggregate costs of production may increase with the entry of a second producer, but the presumption in favour of competition suggests that it is likely to be more than offset by the increase in aggregate benefits.

Where technical advance and product innovation are important, the most important contribution of competition rests on its dynamic efficiency. The existence of a profit-maximising competitor propels the other firm to operate more efficiently, to cater better to its customers and to be more innovative. Few commentators question the importance of competition in this respect. Some doubt the desirability of having a second producer in an industry with significant scale economies but admit the desirability of competition in reducing costs and providing better products through technical advance and innovation. The argument that monopoly may be better than competition among numerous small firms with respect to research and development has also been advanced. However, most researchers argue that oligopolistic competition among big firms is superior to monopoly with respect to innovation. (See Kamien and Schwartz 1975 for a survey.)

The analysis of the preceding section casts doubt even on the desirability of
Should a 'Natural Monopolist' Be Subject to Competition?

Protecting a natural monopoly on grounds of static scale economies. However, even accepting the static inefficiency of having a second producer, the dynamic efficiency so gained may more than offset the loss, as shown in Section 5.

3. The Political Economy of Australian Telecommunications

3.1 Some Specific Aspects of the Telecommunications Industry

In Australia, Telecom was established as a public monopoly in telecommunications. However, in recent years, changes have been made to move towards liberalisation. 'A new framework' was announced in May 1988. The objectives are:

(a) to ensure universal access to standard telephone services throughout Australia on an equitable basis and at affordable prices, in recognition of the social importance of these services;

(b) to maximise the efficiency of the publicly owned telecommunications enterprises – Telecom, OTC, Austsat – in meeting their objectives, including the fulfilment of specific community service obligations and the generation of appropriate returns on investment;

(c) to ensure the highest possible levels of accountability and responsiveness to customer and community needs on the part of the telecommunications enterprises;

(d) to provide the capacity to achieve optimal rates of expansion and modernisation of the telecommunications system, including the introduction of new and diverse services;

(e) to enable all elements of the Australian telecommunications industry (manufacturing, services, information provision) to participate effectively in the rapidly growing Australian and world telecommunications markets; and

(f) to promote the development of other sectors of the economy through the commercial provision of a full range of modern telecommunications services at the lowest possible prices.

We shall refer back to these objectives. There are some characteristics of the telecommunications industry that are relevant to the pricing, regulation, and other issues in this industry, as outlined below.

(i) Telecommunications is characterised by very high overhead costs. This may give rise to scale economies and justify treating it as a case of natural monopoly.

(ii) There have been rapid technological advances in the provision of telecommunications services. This is expected to continue to be the case in the foreseeable future.

(iii) The standard telephone service, a major part of telecommunications, is regarded as an essential service which should be made accessible to everyone. (See point (a) above.) This underlies the practice of cross subsidisation.

(iv) External benefits exist as the caller benefits from being able to reach more people and the receiver may also benefit from a phone call.

(v) There exists a dual pricing system of access rentals and usage charges.

These issues are discussed below.

3.2 Is the Telecommunications Industry a Natural Monopoly?

While the provision of telecommunications services is characterised by very large overheads, their significance can be overstated. First, at any point in time, the total production costs are virtually independent of the amount of usage. Whether more or less phone calls are made in any particular day does not appreciably affect Telecom's production costs. In this sense, there are very low marginal costs of production. However, this does not mean that the short-run marginal cost is low, because economic costs are measured by the best opportunity forgone. In the short run, with given capacity and costs of production, the marginal cost of usage by any consumer is the probable denial of usage by other consumers. When demand is high relative to capacity, the marginal opportunity costs of usage far exceed the marginal costs of production. Hence, despite a huge overhead, marginal-cost pricing need not necessarily result in losses, as frequently believed: see Ng (1987).

Secondly, while the marginal production cost is well below the average cost in the short run, it is much higher in the long run. For the issue of whether to allow entry, it is the long-run costs that are relevant. Though scale economies may still be relevant even in the long run, their magnitude is certainly much smaller than might be suggested by the short-run perspective.
Thirdly, when the importance of technological advance and product innovation is taken into account, the telecommunications industry may no longer be regarded as a natural monopoly despite substantial static scale economies; see also Sharkey (1982, p. 213).

3.3 The Rationality of Universal Access

The first objective of the 'new framework' is to 'ensure universal access to standard telephone services throughout Australia on an equitable basis and at affordable prices, in recognition of the social importance of these services', as mentioned earlier. This objective has no economic basis and is mainly due to political factors partly based on the importance of country voters and the lack of awareness of the effect on the prices charged to urban voters.

In practice, the pursuit of the objectives of universal access and 'community service obligations' is reflected mainly in:

(i) the pricing of access rental well below cost;

(ii) the uniform pricing of access rental in different geographical localities despite the big differences in costs, for example, it is much more costly to supply the same service to a remote locality with very few residents; and

(iii) the pricing of usage charges (especially for trunk calls) well in excess of the cost to finance the required cross subsidy.

While the telephone is an important service, this is no ground for pricing it below costs in order to have universal access. Other goods are not charged the same price irrespective of location.

3.4 External Benefits and the Dual Pricing Structure

Albon (1988) estimated the efficiency loss of the Australian telecommunications pricing structure to be $240 million in 1985–86. Assuming that this remains unchanged relative to the total revenue of Telecom, the loss must now be approximately $440 million per annum. Moreover, this does not include other efficiency losses such as higher operating costs and forgone opportunities.

A consideration (recognised by Albon) which may justify pricing the access rental of telephones below costs is the external benefits referred to above. Two different external benefits are involved. First, as more people are connected to the phone network, the benefit to any subscriber increases. This external benefit justifies pricing access rentals below the costs of providing access. However, this external benefit may be overwhelmed by the consideration of the optimal dual pricing structure discussed below. (Perl 1986 estimated the external benefit for the United States in 1983 at US$4 per subscriber per month. This is in the order of about A$100 per subscriber per annum at current prices.)

Secondly, external benefits may arise if the receiver of a phone call also benefits from the communication since only the caller has to pay the usage charge. However, this benefit should be offset by the cost a receiver may endure. It is not clear whether, on average but at the margin, the receiver of a call views it as a benefit or as a cost. Presumably, the willingness to install a telephone implies that the receiver's expected benefit is positive, but this may largely be accounted for by a small number of important calls which will be made even if the usage charges are much higher. For marginal calls, the receiver may endure a cost rather than a benefit. Even if the relevant magnitude at the margin is positive, this second source of external benefits does not justify charging access rental below cost but justifies charging usage below cost.

While the first source of external benefits may justify pricing access rental below cost, it does not justify favourable treatment of country access since telephones in metropolitan areas also have external benefits. In fact, the external benefits of the latter are likely to be bigger since a telephone in a metropolitan area is reachable by more people at low usage costs. It is true that the alternatives to telephone (for example, visiting) are less costly to people in the same locality. It is more costly to visit a friend in a remote country town. But it is also more costly to visit a friend in another metropolitan area. A telephone in a metropolitan area is not significantly less likely to be contacted by long distance calls than a telephone in a country area. Moreover, the people whom one is likely to call are mostly in the same locality. Hence, the external benefits of a telephone in a metropolitan area are still likely to be bigger when everything has been taken into account. Since the argument of external benefits does not justify pricing access rentals below costs by a larger amount for country people than city-dwellers, such a practice has to be justified by different arguments which are beyond the scope of this article.

The pricing of telephone services is characterised by the existence of both access rentals and usage charges. Most analysts treat these as two separate goods and apply the Ramsey rule for the case of independent demands with the following inverse elasticity rule for social
Should a 'Natural Monopolist' Be Subject to Competition?

optimal pricing under a budget constraint:

\[
\frac{(P_a - MC_a)/P_a}{(P_u - MC_u)/P_u} = E_u/E_a
\]

where \( P \) = price, \( MC \) = marginal cost, \( E \) = elasticity of demand, \( u \) = usage charge, and \( a \) = access rental. In other words, the percentage mark-ups over marginal costs should be inversely proportional to the elasticity of demand.

However, access demand and usage demand are interrelated. The social optimal two-part pricing subject to a budget constraint was first analysed by Ng and Weisser (1974). (Oi 1971 analysed the private optimal two-part pricing of a monopolist.) Though the inverse elasticity rule no longer holds under plausible assumptions, the proportion of 'required revenue' financed by the excess of usage charge over marginal cost increases with the absolute elasticity of the number of consumers and decreases with the absolute income-compensated price elasticity of consumption. ('Required revenue' equals the loss from marginal-cost pricing plus any excess profit that must be made.)

It has been estimated that access demand is typically very inelastic, with the absolute elasticity ranging from zero to 0.1. (See Beggs 1981, Perl 1986 and Taylor 1980.) This suggests fairly high access rentals. This consideration may thus offset the external benefits of access. A more detailed analysis is clearly required in this respect to achieve pricing efficiency.

When the relevant policy-makers decided to price access well below costs to achieve close to universal access, they probably had in mind some equity considerations based on needs. The importance in terms of needs of a standard telephone was regarded as more than reflected by the willingness to pay by low-income groups. This intuition ignores the argument for treating a dollar as a dollar despite the distributional consideration (Ng 1984). Nevertheless, if the pursuit of equality is more effective through the pricing method than the taxation system due, for example, to tax evasion, some justification for the apparently inefficient pricing structure may be provided. (See Kelsey 1988 and Ng 1988.)

4. Cellular Mobile Telephone Service (CMTS)

4.1 The Introduction of CMTS in Australia

There are three components in a cellular mobile telephone (also called cellular radiotelephone) system:

(i) the phone unit which is either installed in a car or as a portable unit;

(ii) cell sites which receive and transmit signals from and to cellular phones; and

(iii) the mobile telephone switching office which links cellular phones to established conventional telephone networks.

<table>
<thead>
<tr>
<th>Date</th>
<th>Number</th>
<th>Percentage growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>13 May 1987</td>
<td>2,449</td>
<td>–</td>
</tr>
<tr>
<td>25 November 1987</td>
<td>16,173</td>
<td>560</td>
</tr>
<tr>
<td>11 May 1988</td>
<td>28,497</td>
<td>76</td>
</tr>
<tr>
<td>23 November 1988</td>
<td>53,997</td>
<td>89</td>
</tr>
<tr>
<td>30 April 1989</td>
<td>80,460</td>
<td>49</td>
</tr>
</tbody>
</table>

While mobile phones were first introduced in 1921, the more powerful version with cellular sites was introduced in Australia in 1987 (it was introduced in the United States in 1983). Soon after its introduction, the number of subscribers grew spectacularly, as in most other countries that have installed the system.

The growth in CMTS in Australia, according to Telecom figures, is reported in Table 1. This rate of increase cannot be expected to persist indefinitely. However, CMTS can be expected to grow strongly, especially if a second operator is permitted to compete. In the near term, CMTS has been forecast to grow at the rate of about 50 per cent per annum, with gradual tapering off in the growth rate over time. Based on these forecasts, together with some plausible assumptions, Section 5 undertakes a simplified cost–benefit analysis of the introduction of a second operator.

4.2 Overseas Experience

The US metropolitan population is divided into 305 markets, each of which is served by two CMTS operators. It is thus a case of duopoly competition. In aggregate, there are now over two million CMTS subscribers in the United States and the number is still growing strongly. Though without some important advantages, the historical division into over 300 markets also creates some administrative problems and inconveniences where intermarket callings and especially intermarket roamings are concerned.

In Europe, there are ten incompatible CMTS systems across different countries. However, from 1991, all countries will convert to a single system called the Pan European Digital Cellular, allowing intercountry callings with CMTS. Most European countries have monopolies in
Should a 'Natural Monopolist' Be Subject to Competition?

the operation of the CMTS system. However, the United Kingdom has a duopoly and France and Germany are going to change over to the UK model.

Reviewing the issue of monopoly versus competition in the telecommunications industry in Britain (including, but not confined to, CMTS), Vickers and Yarrow conclude:

The natural monopoly argument has at least three weaknesses. The first is that it is far from clear that cost conditions are naturally monopolistic with today's technologies. Secondly, if natural monopoly conditions do in fact exist, the removal of barriers to entry would probably not be followed by actual entry, but it would have desirable incentive effects on the incumbent firm or firms. This provides an argument for liberalization when there is uncertainty about whether natural monopoly conditions exist. Thirdly, freedom of entry generally has beneficial effects on the behavior of incumbent firms. As well as allowing the possibility of competition in the product market, it sharpens incentives for internal efficiency, and it tends to undermine the chances of collusion between incumbent firms.

[1988, p. 239]

Discussing the case of US telecommunications, Wenders (1987, p. 206) concludes in favour of competition. For the UK duopoly in CMTS, both the first and second operators spoke favourably about competition. As reported in Telephony 25 January 1988, Racal-Vodafone's Managing Director, Chris Gent, has no doubt on the superiority of duopoly competition over monopoly, citing the resulting advantages of customer choice, greater public awareness, and faster rollout of services. 'Cellular services are cheaper to the end user than they would have been in a monopoly'. What is more remarkable is that even Cellnet (of BT) Sales and Marketing Director, Peter Waller, said: 'The establishment of two operators for cellular communications in the U.K. has worked very well, both for customers, by providing greater variety and keeping prices down, and for the industry by providing a choice so that dealers and retailers can decide which system suits their own business requirements best'.

4.3 Efficient Use of the Radio Frequency Spectrum

Since CMTS is a radiotelephone, its operation involves the use of parts of the radio frequency spectrum. This spectrum has its many competing uses and should be regarded as a scarce resource. A question thus arises as to the efficient allocation of the spectrum. In Australia, a band in the radio spectrum has been reserved for CMTS usage. This band is sufficient to accommodate two operators of CMTS systems. However, any given operator will find it less costly to operate over the whole band than to operate only over half of it. On the other hand, those in favour of competition argue that competition would lead to more subscribers being served within the same band. The users of the spectrum should be charged a price determined by its scarcity value. Ideally, the whole radio spectrum should be allocated according to this principle. There might be certain public-good demands on the use of the spectrum but these could be funded through the public expenditure side. If bands have to be reserved for certain public uses, the rest could be allocated through the market principle. Assuming sufficient numbers of users of the spectrum, the scarcity price could be established through competitive bidding.

However, the radio spectrum has largely been preallocated. Within the band reserved for CMTS usage, the number of competing users may not be large enough to ensure that auctioning will provide the most efficient outcome. In fact, in this setting of a restricted band, it can be shown that:

(i) auctioning the usage of the band is better than maintaining a monopolised usage of it; and

(ii) auctioning need not be the most efficient outcome.

Point (i) is true because if the monopolised usage of the band is a better outcome (due, for example, to scale economies), then the monopolist should be able to bid the highest price for its continued usage of the whole band. Point (ii) is true because even if a monopolist is able to bid the highest price for the use of the whole band, this need not be a most efficient outcome. A monopolist may be able to bid the highest price because of its supernormal profit due to its monopolistic restriction of output. This is illustrated in Figure 4, with AC$^1$ and AC$^2$.

![Figure 4](image_url)
denoting the average cost of supplying the respective aggregate output levels for the case of one and two firms respectively. The most efficient outcome (subject to a breakeven constraint) is for two firms to operate, both selling at the price \( p^2 \). Yet a single operator may be able to bid the highest price for the whole band since this ensures its monopoly power and allows it to capture the supernormal profits of \( p^2 p^m AB \); that is, the shaded area in Figure 4. With only a small preassigned band, whoever controls the whole band becomes an entrenched monopolist.

4.4 The Impact of a Second CMTS Operator

Accepting the objective of meeting community service obligations as given, this section considers whether a second operator in the CMTS system will significantly affect Telecom's capacity to meet those obligations.

First, as recognised by the Australian Telecommunications Corporation Act (Section 27), CMTS is not regarded as part of the standard telephone services where some proximity to universal access may be deemed desirable. Hence, the question of allowing Telecom to remain a monopolist in the CMTS system on this ground does not occur. However, the question remains as to whether Telecom should be allowed to extract its monopoly profits in the cellular area so as to allow it to meet more easily its community service obligations in the standard telephone areas.

The revenue from CMTS remains a small fraction of Telecom's total revenue. If the profits from its CMTS operation are to significantly increase Telecom's capacity to cross subsidise its main operation, the costs of distortion in the CMTS area will be unattractively large. On a static consideration of second-best pricing, it remains possible that some contribution from its CMTS operation is desirable. However, if the more important dynamic benefits of competition are to be fully realised, it is important to keep sectors subject to competition free of positive or negative cross subsidisation.

Moreover, it is questionable whether a second operator of the CMTS system will adversely affect Telecom's capacity. Although Telecom's revenue from its CMTS operation may be less than it would be without a second operator, it may be increased from the contributions of the second operator to Telecom's public switched telephone network (PSTN) through which most communications have to be connected. Given that a second operator will increase the aggregate CMTS market, the net effect on Telecom's profits may not be negative even before we take into account the likely higher efficiency following competition.

5. Competition versus Monopoly in CMTS: An Indicative Cost–Benefit Analysis

The discussion above provides a strong case in favour of competition. However, it falls short of providing an indication of the likely magnitude of the gain in allowing competition in CMTS. This section indicates the likely order of magnitude involved under various assumptions about the relevant parameters, rather than providing precise figures. While the degree of precision may be improved by a more detailed analysis, some degree of imprecision and uncertainty always remains.

It is important that the methodology used is valid and that the assumptions that are made are clarified. If a wrong framework is used, one will get a wrong result no matter how precise the estimates of the relevant parameters. The method used here is based on the estimation of changes in consumers' surplus which has a long tradition in economics and is widely used in cost–benefit analyses. Though not without some conceptual ambiguities, its use can be defended (Ng 1983, ch. 4).

5.1 The Framework of Analysis

A simple partial-equilibrium framework is adopted for estimating consumers' surplus. It is recognised that, when interactions with the rest of the economy with second-best complications are taken into account, the partial-equilibrium measures may either underestimate or overestimate the actual gains or losses. However, since CMTS is a relatively small sector of the whole economy, the partial-equilibrium approach is an acceptable approximation. To avoid overestimating the benefits of competition, we adopt a downward adjustment factor.

Start with the demand function for CMTS. For simplicity, the quantity axis measures only the number of cellular phones demanded. This is thus an access demand, holding usage charges at some given figure and with the total usage charges included in the access price. As mentioned in Subsection 3.4, access demand for standard phones is very inelastic, with absolute elasticity less than 0.1. The access demand for cellular phones (with usage charges included) may be expected to be much more elastic, probably not less than 0.5 in the long run. However to ensure no overestimation of the benefit of competition, an even higher figure of one is used. (The lower the absolute elasticity, the higher the benefits of competition, holding other variables in the following
should a ‘natural monopolist’ be subject to competition?

exercise unchanged.) This elasticity of unity is taken as applicable throughout the whole range. Hence, the demand function may be described as

\[ Q = \frac{A}{P} \]  

(1)

where \( Q \) = quantity, \( P \) = price, and \( A \) is a parameter denoting the magnitude of the market demand. It can be expected to increase over time.

For simplicity, differential usage rates by different consumers are ignored as are differences in the quality of the cellular phones used. The possible difference between renting and buying a cellular phone is also ignored. Hence each consumer is charged the same price (inclusive of annual rental plus a standard usage charge) and it costs the CMTS operator the same amount. The area above the price line but below the demand curve measures the consumers’ surplus.

Consumers’ benefit is the area under the demand curve. However, with \( P = NQ \) from (1), the height of the demand curve approaches infinity as \( Q \) approaches zero. This problem is avoided by measuring consumers’ benefit between \( Q \) and \( Q \), where \( Q \) is some small value of \( Q \). Then, the consumers’ benefit is measured by

\[ \int_{Q}^{Q} A dQ = A (\ln Q - \ln Q) \]  

(2)

Over time, the demand curve shifts outward due to higher incomes, a bigger population and greater awareness of CMTS. For simplicity, we adopt a constant rate of growth \( G \). It is recognised that, for the case of CMTS, there is more likely to be a decreasing growth rate over time. However, if a constant rate of growth \( G \) below the initial growth rate is selected, this may approximately reflect the case of a decreasing growth rate, taking account of a discount factor. While the divergence between the simulated growth path and the actual growth path eventually diverges to a large figure in the very distant future, the estimation in terms of present value may yet be accurate due to time discounting. In any case, after presenting the results for estimation from the present to the infinite future, we also provide the results for the first twenty years only.

As illustrated in Figure 5, if the (real) price decreases over time, there will be a further increase in quantity demanded (from \( E' \) to \( E_2 \)) apart from the rightward shift in the demand curve.

To model the effects of the introduction of competition, it is assumed that the price charged by both duopolists is the same and also the same as what would be the case under monopoly. This is likely if Telecom persists with whatever is its pricing policy in the absence of competition and AUSTEL performs its function to ensure the absence of collusion. Competition may lead to better service, higher quality and even lower prices. However, we may lump all these higher benefits into consumers’ benefit per unit of product defined to have a similar price as under monopoly. Then the effects on consumers are all captured by the higher demand under competition. This increase in consumers’ surplus is indicated by the shaded area in Figure 6 where the superscripts m and c stand for monopoly and competition respectively.

The comparison illustrated in Figure 6 is static. Over time, both \( D^m \) and \( D^c \) shift rightward. The following allows for these shifts and measures the gain to consumers due to the introduction of competition. This gain is then compared with the possible losses to Telecom due to the introduction of competition.

It may be thought that, by having a higher demand curve for the case of competition, we are prejudging the case. However, there is almost universal agreement that competition in CMTS will bring more benefits to consumers through ‘increasing customer responsiveness and faster technological innovation’ (Evans 1988, p. 49). If prices fall per quality-adjusted unit, increases in consumers’ surplus are assured.
5.2 First Estimation: Infinite Time Horizon

In the absence of competition, let the demand curve shift out at a constant rate \(G\), as indicated by

\[
A_t^m = A^m e^{G t}
\]  

where \(A_t^m\) is the demand parameter for the case of monopoly at time \(t\) and \(A^m\) is its initial value at \(t = 0\). For the purpose of comparison, this is the time upon the introduction of competition.

Let the price (always in real terms) fall at a constant rate \(p\),

\[
P_t = P e^{-pt}
\]

There is no need to use the \(m\) superscript for \(P\), since \(P(t = \infty) = P_c\).

From (1), (3) and (4) we have

\[
Q_t = \frac{A^m e^{G t}}{P} = a e^{G t}
\]

where \(a = A^m / P\), \(g = G + p\).

At any point in time \(t\), consumers' surplus under monopoly is given by

\[
C^m \mid t = \int_0^t \left[ \frac{A e^{G t}}{P} \ln \left( \frac{A e^{G t}}{P} \right) - \ln Q_t - \frac{(Q_t - Q) P_t}{P} \right] dt
\]

The present value (that is, at \(t = 0\)) of consumers' surplus from time zero to infinity, at a real discount rate of \(r\), is given by

\[
C^m \mid 0 = \int_0^\infty \left[ \frac{A e^{G t}}{P} \ln \left( \frac{A e^{G t}}{P} \right) - \ln Q_t - \frac{(Q_t - Q) P_t}{P} \right] e^{-rt} dt
\]

Consider now the case of competition. At time \(t = 0\), the magnitude of demand is taken as the same as the case under monopoly; that is, \(A_0^c = A^m\). However, \(A_t^c\) grows faster than \(A_t^m\).

\[
A_t^c = A^c e^{G' t}
\]

where \(G' > G\).

With (8) replacing (3) but with everything else similar, the present value of consumers' surplus with competition is

\[
C^c \mid 0 = \int_0^\infty \left[ \frac{A^c}{P} \ln \left( \frac{A^c}{P} \right) - \ln (Q_t^c - Q) - \frac{(Q_t^c - Q) P_t}{P} \right] e^{-rt} dt
\]

It can be shown that the gain expressed above decreases with \(A\), \(Q\), \(P\) and \(r\), and increases with \(G\), \(\Delta G\) and \(p\). For low estimates of the gain from competition, select low values for those parameters having a positive effect on the gain. The values of \(A\) and \(P\) are determined by the situation when competition is being introduced. Take \(P = 1500\), \(A = 150,000,000\). This implies an initial number of 100,000 subscribers, since \(Q_0 = A/P\). Selection of values for other parameters is more controversial. The following choice deliberately biases towards a low estimate of the gain:

\[
r = 0.12, \ P = 0, \ G = g = 0.08, \ \Delta G = 0.005 \text{ and } G' = 0.085.
\]

It may be thought that \(r = 0.12\) is a low value for the discount rate. However, realising that the figures are in real terms (that is, dollar figures in the future have not included inflationary effects), a real discount rate of 12 per cent really means a nominal discount rate in excess of 20 per cent, given that the inflation rate is about 7.5 per cent (1.12 x 1.075 = 1.204). This 20 per cent is quite high even in comparison with the current high rates of interest which are expected to fall in the near future. A nominal discount rate of 20 per cent means that a (nominal) dollar in twenty years time is valued at 1.15 cents now. A real discount rate of 12 per cent means that a (real) dollar in twenty years time is valued at 7.76 cents now. This is a rather high rate of discount.

Demand is assumed to grow at 8 per cent per annum under monopoly and is 0.5 per cent higher under competition, a fairly conservative figure. Prices (real) are assumed unchanged (\(p = 0\)). If they fall (\(p > 0\)), the estimated gain will be higher.

With these values the first estimate of the gain to consumers is

\[
\text{Gain} = \$4839500000
\]

This figure of \$4.84 billion is, in per capita terms, about \$280. At a discount rate of 12 per cent, it is equivalent to \$580 million (or about \$34 per head) in real terms every year. Or, at a nominal discount rate of 20 per cent, it is equivalent to \$968 million (or about \$56 per head) in dollar terms every year. This is a high but not an implausible figure.

Due to the partial-equilibrium framework, some adjustments may be necessary. Also, it might be doubted on other grounds that all the higher demand...
under competition could be regarded as gain. For these reasons, we may wish to adopt an adjustment factor to ensure that the gain from competition is not overestimated. Even if we adopt an adjustment factor as high as 50 per cent, we still have

Adjusted Gain = $2,420,000,000 (12)

which is equivalent to $290 million every year in real terms.

On the producer side, if a second operator voluntarily enters, the only possible loser is Telecom. It is not easy to estimate the gain or loss to Telecom, but a loss of $290 million in real terms every year is out of the question. Thus, competition passes the test of cost–benefit analysis provided only that prices are not increased after the entry of a second operator (as might be the case if there is collusion between the duopolists). If a recommendation made in the next section is adopted, this collusion can almost certainly be avoided.

The estimate above for the gain to consumers in a framework of an infinite horizon may disturb some people since the postulated trend cannot be relied upon to apply indefinitely into the future. Therefore, an estimate for a time horizon of twenty years is presented below.

5.3 Second Estimation: Finite Time Horizon

If the time horizon is finite at \( t = T \), we have, from the second last equation in (7),

\[
CS^c_{T,0} = \frac{A}{r - G'} \left[ \left( \frac{g'}{r - G'} - \ln Q - 1 \right) \left( 1 - e^{(G' - r)T} \right) \right] + \frac{PQ}{\rho + r} \left( 1 - e^{(\rho + r)T} \right)
\]

(13)

For the case of competition,

\[
CS^c_{T,0} = \frac{A}{r - G'} \left[ \left( \frac{g'}{r - G'} - \ln Q - 1 \right) \left( 1 - e^{(G' - r)T} \right) \right] + \frac{PQ}{\rho + r} \left( 1 - e^{(\rho + r)T} \right)
\]

(14)

The gain to consumers from competition is thus

\[
CS^c_{T,0} - CS^m_{T,0} = \frac{A}{r - G'} \left[ \left( \frac{g'}{r - G'} - \ln Q - 1 \right) \left( 1 - e^{(G' - r)T} \right) \right] - \frac{A}{r - G} \left[ \left( \frac{g}{r - G} - \ln Q - 1 \right) \left( 1 - e^{(G - r)T} \right) \right] + \frac{PQ}{\rho + r} \left( 1 - e^{(\rho + r)T} \right)
\]

(15)

If it is conservatively assumed that the demand for CMTS is increasing at \( t = 0 \) (introduction of competition) at 30 per cent per annum and that this rate of growth falls by 10 per cent each year thereafter (that is, 27 per cent in the next year, 24.3 per cent the year after next etc.), the rate of growth will fall to 3.65 per cent after twenty years (that is, at \( t = T = 20 \)). Since a growth rate of 3.65 per cent can easily be sustained just by population and income growth, this is not an unreasonable assumption. The path of the rate of growth over time is illustrated in Figure 7. This path of growth may be approximated by a constant growth rate of 15 per cent. This is a low approximation since, at a discount rate in excess of 10 per cent, earlier figures have a much higher impact on the result.

Take the growth rate in demand under competition to be three-quarters of a per cent higher than in the case of monopoly. Other parameters are taken as the same as in the previous estimation. Substituting \( \rho = 0.12, G = g = 0.15, G' = g' = 0.1575, Q = 1000, A = 150,000,000 \) and \( P = $1500 \) into (15) gives an estimate of the gain to consumers from competition in the first twenty years as (in present value at \( t = 0 \))

\[
Gain = $2,400,000,000 (16)
\]

Applying an adjustment factor of 50 per cent, the resulting adjusted gain is still in excess of $1 billion. Again, no reasonable estimate on the loss, if any, of the existing monopolist can be expected to exceed this sum.

6. Recommendations

From the above analysis, we make the following recommendations.

(i) The general case in favour of competition is overwhelming. While the
argument of natural monopoly has been used to justify the disallowance of competition, telecommunications may not be regarded as a natural monopoly when account is taken of the expanding boundary due to rapid product innovation and technical advance. Moreover, if it were a natural monopoly, it would need no protection. There being no valid grounds justifying the monopoly of CMTS, we recommend that competition be permitted.

(ii) Due to the rapid growth in demand in CMTS, the benefits of competition would be greater the earlier competition is introduced.

(iii) Within the existing band of the radio frequency spectrum allocated to CMTS, more than two operators does not seem to be feasible. In the absence of an increase in this band, a duopoly situation may result after the introduction of competition, with no possibility of further entries. Should the duopolists collude between them to exploit consumers, the gain from competition may be largely offset. This remote possibility could be avoided by ensuring that Telecom, a public organisation, does not follow collusive price increases. If Telecom commits itself to a low price policy, inefficient entry can be avoided. An operator willing to enter despite Telecom’s commitment can be expected to be efficient.

(iv) Given that competition is permitted, Telecom should review its recently announced plan to spend heavily on capital investment in CMTS. Such a policy may make good sense in the absence of a second operator or as an entry prevention device. However, its scope needs to be reviewed given the entry of a second operator.

(v) At a more general level, the objective of community service obligations and the related Telecom pricing structure should be reconsidered in view of the allegation of huge efficiency losses and in view of the argument of treating a dollar as a dollar. While equity issues have to be taken into account, they have to be considered simultaneously with efficiency issues.

(vi) The government should reconsider the allocation of the radio frequency spectrum, taking account of national objectives and economic efficiency.

First version received October 1989; final version accepted June 1990 (Ed.).

References
Beggs, J. J. 1981, 'The demand for telephone services in Australia and the welfare implications of alternative pricing policies', mimeo, Australian National University, Canberra.
Should a 'Natural Monopolist'
Be Subject to Competition?